开源组件在新华社数据服务平台中的应用

摘　要：随着媒体融合发展、移动互联网的高速发展，传统的 IT 架构不能适应新形式下的媒体报道要求。为适应新要求，新华社推进了四大应用平台建设。数据服务平台作为新华社数据核心，为四大应用平台提供了底层数据支撑。数据服务平台采用了微服务架构，使用大量的开源、分布式组件。本文对数据服务平台及其使用的组件进行了介绍，并总结了数据服务平台带来的数据管理两大变化。

关键词：数据服务平台；大数据；开源组件
中图分类号：TQ018　　文献标识码：A
文章编号：1671-0134（2018）07-065-04　　DOI：10.19483/j.cnki.11-4653/n.2018.07.018

文 / 乔爱军

随着媒体融合发展的要求，移动互联网的深入普及，新华社传统的 IT 架构已经不能适应新时代报道的需求。新华社于 2016 年底开始构建新一代技术体系，目前，基于开源技术的全媒体采编发平台（新采编发）、全媒体供稿平台（新供稿）、办公协同平台（新 OA）、全媒体业务管理平台（新闻热点、新闻线索、落地统计、报道指挥）四大应用平台基本建设完成。数据服务平台作为新华社数据汇聚中心，为上述四大平台提供底层支撑。现数据服务平台已经初步建成，提供了数据接入、存储、服务、处理、分析、统计、应用等服务，实现了社内稿件、引进资源、互联网数据、用户信息、行为数据和服务审计数据的聚合，为各应用系统提供数据服务。数据服务平台采用分布式、开源的架构体系，辅助于一体化分层的监控体系，为新华社融合报道提供了强有力的支撑。

1. 数据服务平台概况

数据服务平台对外提供四大类的服务，即服务管理、数据服务、大数据计算服务和 GlusterFS 文件存储服务。服务管理实现服务接口的统一管理，提供服务注册、发现、路由、管理等功能。服务管理除了注册和管理数据服务平台自己的服务外，还注册了用户认证管理系统和办公协同平台等其他系统的服务。数据服务目前提供了基础服务、全文检索、数据订阅、语义分析、推荐、标签、用户信息、资源管理 8 类 99 个服务。大数据计算服务在新华社总社、东坝机房分别部署了 17 个节点和 14 个节点。大数据计算服务可为全社提供 Spark、Storm 计算服务，HDFS、HBase、Hive 存储分析能力。GlusterFS 分布式文件存储服务，同样分为两个集群，目前总社主集群 30 个节点，东坝备份集群 12 个节点，已经为四大平台和数据服务平台提供了 322T 的空间。

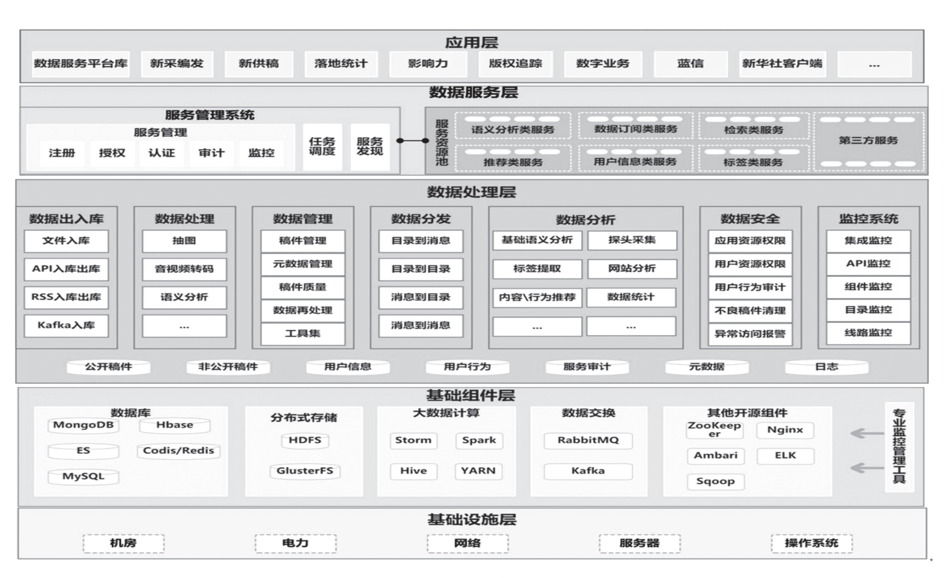


图 1　数据服务平台总体架构图

数据服务平台自下而上分为四层，即基础组件层、数据处理层、数据服务层和应用层，依托于基础设施层。

2. 基础组件层

基础组件层使用的主要开源组件有 18 个，为数据服务平台、全媒体采编发、新 OA、落地统计、影响力分析和项目管理系统提供服务。根据开源组件在数据服务平台中的功能，我们将其分为数据库、分布式文件存储、大数据计算、数据交换、其他五类。

2.1 数据库类组件

数据库类组件主要包括 MongoDB、ES、MySQL（MariaDB）、HBase、Codis/Redis。MongoDB 用于存储社内稿件，HBASE 和 ES 除存储社内稿件外还存储互联网数据，MySQL 用于系统运行所需要的关系型数据。Codis/Redis 在数据服务平台中存储频繁使用的业务数据，如授权使用的分类、用户行为数据、稿件推荐数据等。

MongoDB 是专为可扩展性、高性能和高可用性而设计的数据库，介于关系数据库和非关系数据库之间的产品。它可以从单服务器部署扩展到大型、复杂的多数据中心架构。利用内存计算的优势，MongoDB 能够提供高性能的数据读写操作。 MongoDB 的副本复制和自动故障转移功能使应用程序具有企业级的可靠性和操作灵活性，分片机制提供了方便的横向扩展能力。其非结构化的特性，方便存储各种类型稿件数据、用户数据、日志数据。在数据服务平台中使用 MongoDB 数据库代替了原来的 Oracle 数据库存储社内稿件，作为核心应用数据库。

ES 是 Elastic Search 的简称，它是由多个 Lucene 实例组成的分布式检索和分析系统，是分布式的搜索引擎和数据分析引擎，可以对海量数据进行近实时处理。ES 结合了全文检索、数据分析以及分布式技术，提供了强大功能。数据服务平台中使用 ES 做检索处理。

MySQL 是一种大家广泛熟悉的关系数据库。但 MySQL 被 Oracle 公司掌控后，MySQL 原来开发者创立了 MariaDB 分支。MariaDB 与 Mysql 高度兼容，并保证开源免费。数据服务平台主要使用 MySQL 保存服务和应用信息、线路和目录监控信息、分发信息等系统元数据、配置等结构化数据。

HBase 是一个分布式的、面向列的开源数据库，该技术来源于 Fay Chang 所撰写的 Google 论文 Bigtable。HBase 不同于一般的关系数据库，它是一个适合于非结构化数据存储的数据库。HBase 主要用于存储社内稿件和互联网稿件等数据，用于离线数据分析。

Codis/Redis 是 KeyValue 型内存数据库。Codis 是一个分布式 Redis 解决方案，通过 Codis Proxy，将多台 Redis 服务器集中起来使用，实现 Redis 服务器横向扩展。对于上层的应用来说，连接到 Codis Proxy 和连接原生的 Redis Server 没有明显的区别，上层应用可以像使用单机的 Redis 一样使用 Codis。 Codis 底层会处理请求的转发，不停机的数据迁移等工作，这些后台事务对于前面的客户端来说是透明的，可以简单地认为后边连接的是一个内存无限大的 Redis 服务。

2.2 分布文件存储类组件

分布式文件存储类组件包含 HDFS、GlusterFS。

HDFS 是 Hadoop 分布式文件系统的简称。HDFS 是一个高度容错性的系统，适合部署在廉价通用的机器上。HDFS 能提供高吞吐量的数据访问，非常适合大规模数据集上的应用。HDFS 放宽了一部分 POSIX 约束，实现流式读取文件系统数据的目的。数据服务平台中，HDFS 主要为 Hive、HBase 等组件提供文件系统支撑服务，存储日志及用户行为数据等。

GlusterFS 是一种分布式文件系统，存储的对象为文件。它将多台服务器上的空间统一管理起来形成存储池供外界使用。相比于其他的分布式文件系统，GlusterFS 最大的优点在于运维简单，极易上手。GlusterFS 的架构为三层：集群、卷、Brick。在数据服务平台中，按照存储文件大小将集群分为三类：一类是小文件集群，存储 1k–100K 左右大小的文件，主要涉及 CNML 文件，图片的缩略图，部分音视频的关键帧；另一类是中文件集群，存储 500k–10M 左右的文件，一般涉及图片、Word、Pdf 等附件；最后一类是大文件集群，存储 10M 以上的文件，主要涉及音视频文件。每个集群会根据业务需要，分别建立不同的卷，每个卷由指定的集群中的部分或全部服务器下的 Brick 组成。Brick 是一个被建立的目录，用来存储数据。

2.3 大数据计算类组件

大数据计算类组件包括 Storm、Hive、Spark、YARN。

Storm 用于分布式的实时流式数据处理。应用场景有实时分析、连续计算、在线学习、分布式 RPC、ETL 等。Storm 集群由 Nimbus、Supervisor 节点组成，Nimbus 是主控节点，用于提交任务、分配集群任务，集群监控；Supervisor 是计算节点，接受 Nimbus 分配的任务，管理属于自己的 Worker 进程；Nimbus 和 Supervisor 通过 Zookeeper 进行协同。数据服务平台中，使用 Storm 实

现数据的格式转换和数据转储，为数据分析提供数据基础。

Hive 是建立在 Hadoop 上的数据仓库基础构架。它提供了一系列的工具，可以用来进行数据提取转化加载（ETL），可以存储、查询和分析存储在 Hadoop 中的大规模数据。Hive 定义了简单的类 SQL 查询语言，称为 HQL，它允许熟悉 SQL 的用户查询数据。同时，这个语言也允许 MapReduce 开发者开发自定义的 mapper 和 reducer 来处理复杂的分析工作。

Spark 是专为大规模数据处理而设计的快速通用的计算引擎。Spark 是加州大学伯克利分校的 AMP 实验室所开源的类 Hadoop MapReduce 的通用并行框架。Spark 拥有 Hadoop MapReduce 所具有的优点，但不同于 MapReduce 的是，Spark 中间输出结果可以保存在内存中，从而不再需要读写 HDFS。因此，Spark 能更好地适用于数据挖掘与机器学习等场景。

YARN 是大数据集群中资源的管理和调度模块。YARN 的基本思想是将资源管理、任务调度和监控分成不同的模块。主要方法是创建一个全局的 ResourceManager（RM）和为每个应用程序创建的 ApplicationMaster（AM）。这里的应用程序是指单一的作业或作业的 DAG（有向无环图）。ResourceManager 控制整个集群并管理基础计算资源的分配。ResourceManager 将各个资源部分（计算、内存、带宽等）精心安排给 NodeManager。ResourceManager 还与 ApplicationMaster 一起分配资源，与 NodeManager 一起启动和监视它们的基础应用程序。

2.4 数据交换类组件

数据服务平台使用了两种消息交换组件，分别是 RabbitMQ 和 Kafka。使用消息交换组件有两大优点：一是系统之间解耦，降低系统间的耦合性；另一个是业务削峰，当生产者大量产生数据时，消费者无法快速消费，消息组件作为中间层来保存这个数据，达到业务削峰的目的。

RabbitMQ 是实现 AMQP（即 Advanced Message Queuing Protocol，高级消息队列协议）的一种消息中间件，最初起源于金融系统，用于在分布式系统中存储转发消息。在数据服务平台中，利用 RabbitMQ 实现各应用程序间的消息传递，完成稿件处理的各流程，是数据服务平台的核心应用。下图为数据服务平台中新闻稿件处理流程图，从图中可以看到核心的消息队列是由 RabbitMQ 来实现的，它将数据服务平台内部各模块和外部各系统连接起来。

Kafka 是一个分布式的流平台，它有三个关键的功能：发布和定阅流式记录，类似消息队列或企业消息系统；以容错的方式存储流式消息记录；及时处理流式消息记录。Kafka 主要用来构建实时的流式数据管道，构建实时流式数据应用。在数据服务平台中，Kafka 主要用于互联网数据、用户行为数据、服务审计数据、日志数据的交换和传输。比如，落地统计系统从互联网采集的数据经过清洗和过滤后，放到 Kafka 的 topic 中，供数据服务平台入库或其他系统进行处理分析。探头系统采集的用户行为数据、服务管理系统的服务审计数据、应用系统和基础组件的日志数据通过 Kafka 传给 Storm 进行处理或存入 HDFS 供其他应用分析处理。

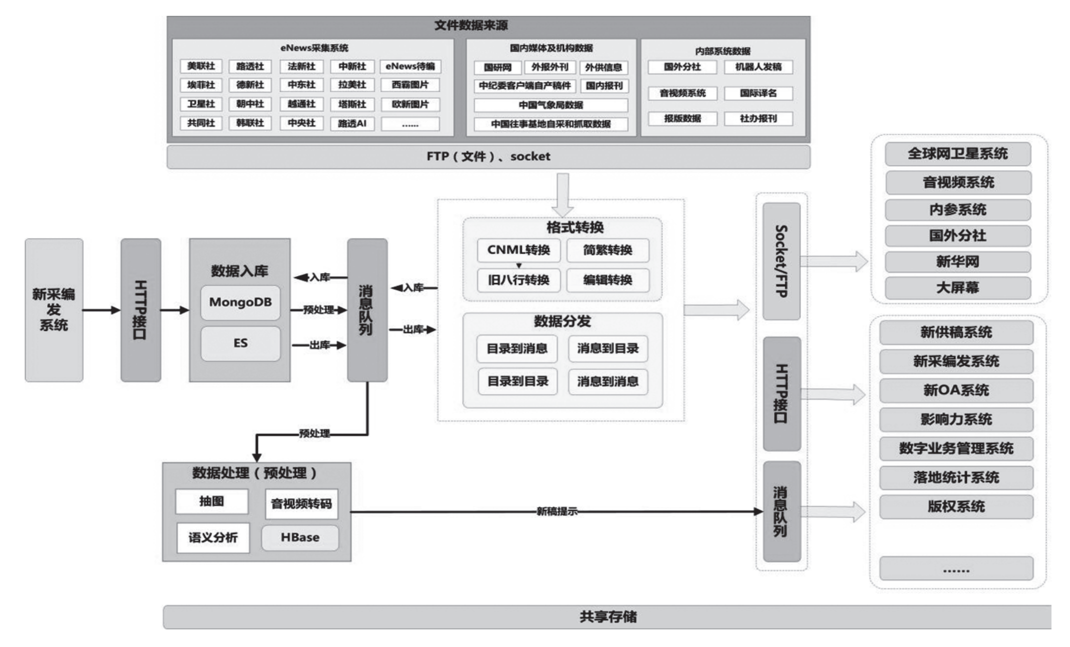


图 2　新闻稿件处理流程

2.5 其他类组件

其他类开源组件包括 Zookeeper、Nginx、Sqoop、Ambari、ELK。

Zookeeper 是为分布式应用提供协同服务的高性能分布式系统，它可以提供配置信息维护服务、命名服务、分布式同步服务、组服务。大部分基于分布式的大数据组件需要 Zookeeper 提供支撑。

Nginx 是俄罗斯人编写的十分轻量级、高性能的 HTTP 和反向代理服务器，也是一个 IMAP/POP3/SMTP 代理服务器。Nginx 以事件驱动的方式编写，具有非常好的性能，同时也是一个非常高效的反向代理、负载均衡。Nginx 具有很高的稳定性，支持热部署，启动容易，并且几乎可以做到 7*24 不间断运行，即使运行数个月也不需要重新启动。还能够在不间断服务的情况下，对软件版本进行升级。数据服务平台中大量应用使用 Nginx 做代理，实现负载均衡，或直接提供 Web 应用服务。

Sqoop 是一个用来将 Hadoop 和关系型数据库中的数据相互转移的工具，可以将一个关系型数据库（如 MySQL ,Oracle ,Postgres 等）中的数据导入进到 Hadoop 的 HDFS 中，也可以将 HDFS 的数据导入到关系型数据库中。对于某些 NoSQL 数据库它也提供了连接器。数据服务平台使用此工具将 MySQL 中的用户数据导入到 HIve 中。

Ambari 是一种基于 Web 的管理工具，支持 Apache Hadoop 集群的部署、管理和监控。Ambari 已支持大多数 Hadoop 组件，包括 HDFS、MapReduce、Hive、Pig、Hbase、Zookeeper、Sqoop 和 Hcatalog 等。使用 Ambari 可以方便部署安装 Hadoop 集群，起停 Hadoop 集群中的相关组件，监控 Hadoop 集群及相关组件的运行状况。

ELK 由三个开源组件（ES、Logstash、Kibana）构成的日志收集、处理和展示工具。数据服务平台使用 ELK 做应用日志分析。

3. 数据服务层和数据处理层

数据服务层和数据处理层的部分功能使用 Spring Cloud Netflix 微服务架构实现。通过 Eureka + Zuul + Ribbon + Feign + Hystrix 构建微服务架构，将数据处理层中的公共模块封装成服务接口，注册到 Eureka 服务器上。Eureka 维护着每个服务的生命周期，并通过心跳确定服务是否正常。Zuul 部署在 Eureka 前端，作为智能路由为外部请求提供统一入口。服务客户端使用 Feign 方式调用，通过 Ribbon 实现服务端的负载均衡；Hystrix 断路器为避免发生雪崩效应而引入，对服务延迟和故障提供更加强大的容错能力。

结语

数据服务平台利用开源组件的分布可扩展性，结合数据服务层使用的微服务架构，使新华社数据管理架构出现两大转变：第一个转变是从库到数据的变化。数据的存储方式由集中式的关系型数据库转换成了分布式数据库；数据加工由原来简单的增、删、改、查转换为数据的处理、分析、挖掘；数据的种类由原来单一的新闻稿件扩展到了互联网数据、用户行为数据、日志数据等多元化数据；数据量也由发生了巨大变化，原来每日生成的稿件数据约几万条，现在每日采集清理后的互联网数据达三百多万条，另外还有大量的用户行为和日志数据。第二个转变是从系统到平台的转变。原来的 IT 系统以单个系统为单位对外提供功能，随着时间的发展，新华社的 IT 系统林立，功能交错依赖，众多新系统上线后，老旧系统不能及时下线并占用大量资源，增大了运维难度；平台的建立，完成了从提供功能向提供服务能力的转变，把功能分解为服务，统一注册管理，避免了重复建设，降低了运维难度；平台的服务对象，由单一应用扩展到同时为多个应用提供服务；平台建设由原来的使用商业化产品转向使用开源组件，节省了系统建设成本，提高了系统上线速度，满足了互联网时代产品快速上线、不断迭代的要求；新的平台使用水平分层、横向扩展的分布式架构，取代原来系统使用的垂直一体架构，从而实现服务能力快速、低成本、动态平滑扩展。媒

参考文献

[1] The Apache Software Foundation. Apache Hadoop. http：// hadoop.apache.org，2018（06）：13.

[2] 王方旭，基于 Spring Cloud 实现业务系统微服务化的设计与实现，电子技术与软件工程，2018（08）.

[3] 王占宏，王战英，顾国强，马国春，吕振华 . 分布式弹性搜索研究与实践，微型电脑应用，2014 （30）：7 .

（作者单位：新华社技术局）